



Industry Trends in Storage

Keith Parris

Systems / Software Engineer

Multivendor Systems Engineering

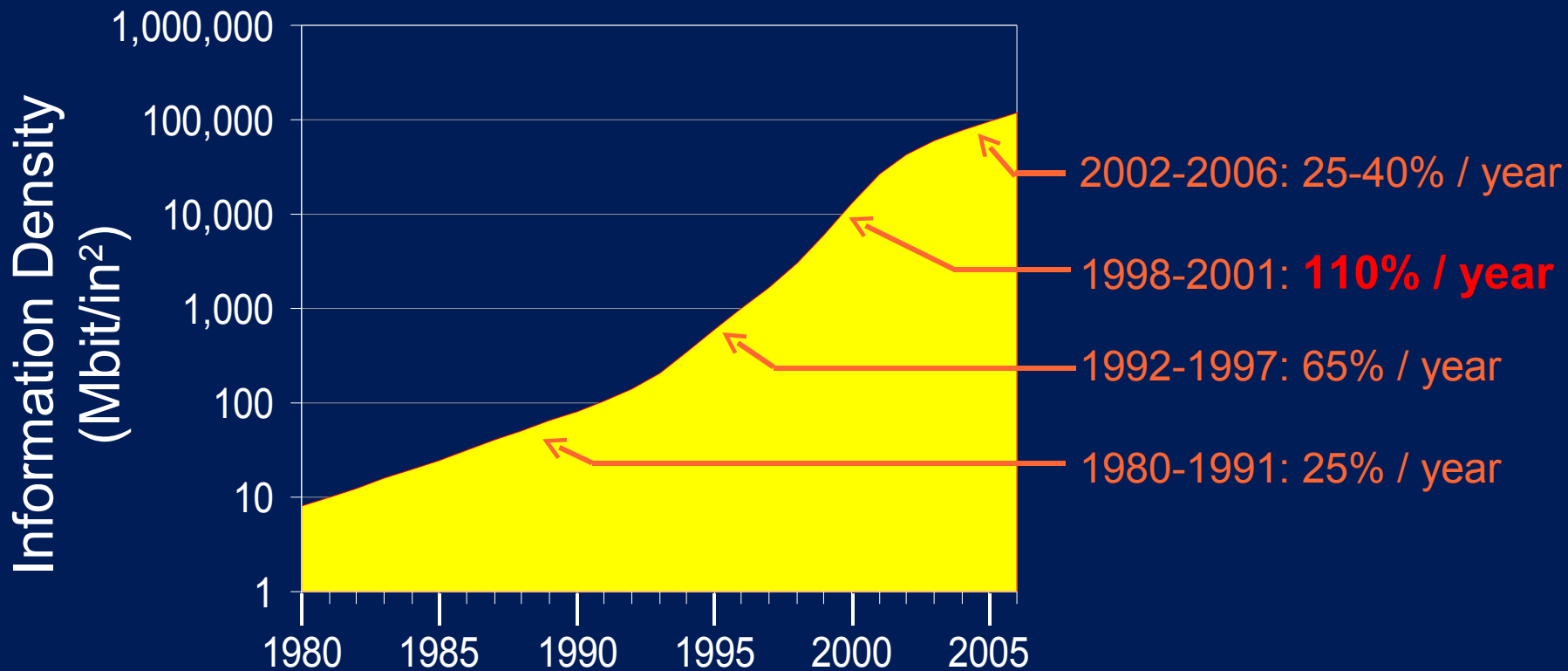
HP Services



Disk Capacity Growth Continues



- But growth rate slows:



Source: Richie Lary, TuteLary, LLC

“The Superparamagnetic Limit has Entered the Building.”

Richie Lary
TuteLary, LLC

Superparamagnetic Limit



- Magnets get unstable if they get too small
- Stability depends on:
 - Magnetic domain volume
 - Magnetic domain shape
 - Magnetic material coercivity
 - Temperature
 - Magnetic polarity of surrounding domains

Superparamagnetic Limit: Workarounds

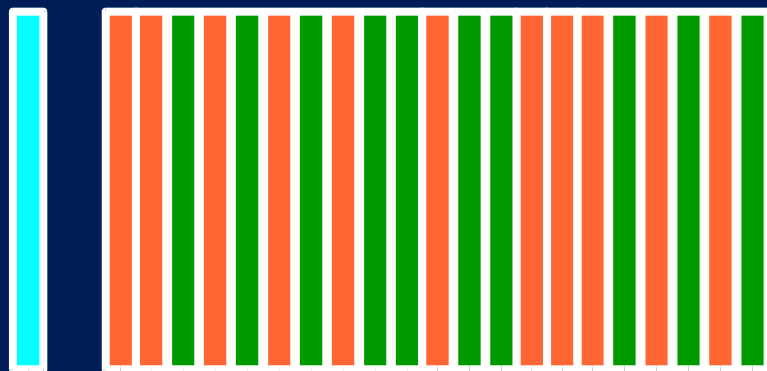


- Workarounds:
 - Make magnetic domains “squarer”,
 - Make magnetic domains “deeper”, or
 - Temporarily modify coercivity

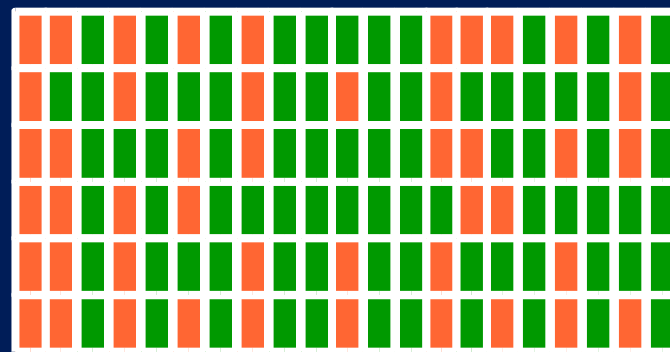
Superparamagnetic Limit: Squarer Bits



- Squarer: Higher track density vs. higher linear density
 - Following track within $\pm 10\%$ of width becomes more difficult
 - Likely to hit lithography limit in 2005 at current rate



Head



Head

Superparamagnetic Limit: Deeper Bits

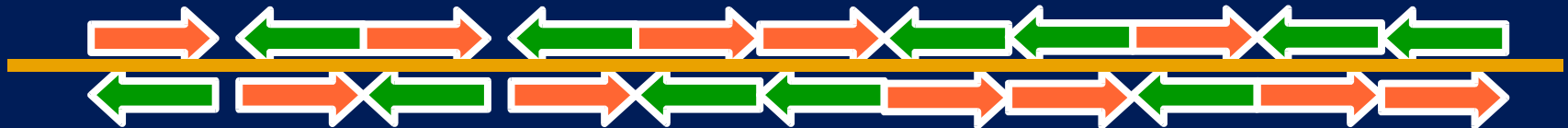


- Deeper:
 - IBM “pixie dust” approach
 - non-magnetic layer between magnetic layers which are magnetized in opposite directions
 - or the elusive Vertical or Perpendicular Recording

Conventional Longitudinal Recording:



Multiple layers with intervening non-magnetic layer:



Vertical Recording:



Superparamagnetic Limit: Modify Coercivity



- Modify Coercivity:
 - Heat Assisted Magnetic Recording (HAMR)
 - Heat high-coercivity magnetic material with laser to “soften” it magnetically just before recording

Growth in Capacity vs. Performance



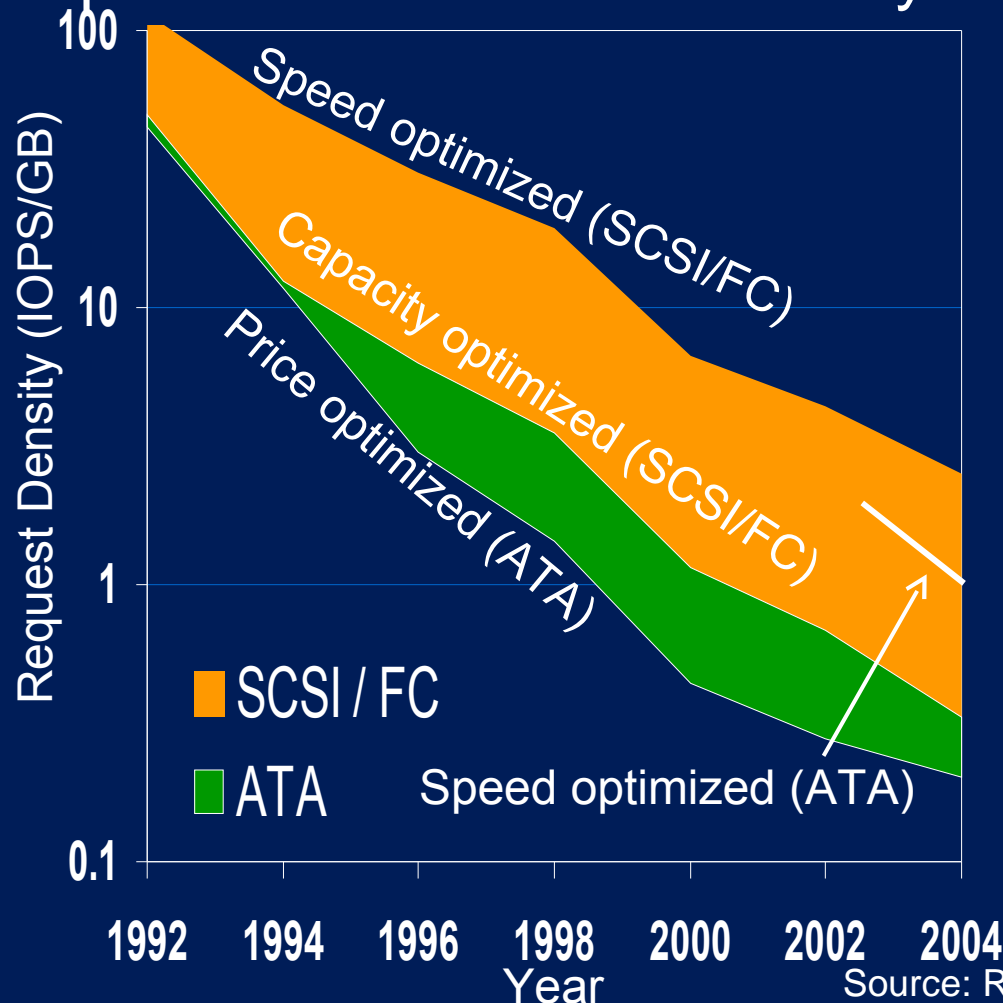
- Progress over the last 30 years:

| | | | |
|----------------------------------|--------|--------|-------|
| Capacity: Data per spindle | 300 MB | 300 GB | 1000x |
| Performance: Random seek time | 30 ms | 3 ms | 10x |

Capacity vs. Performance



- Performance per unit of data has actually fallen:



Source: Richie Lary, TuteLary, LLC

Disk Market Segmentation



- Two classes of products:
 - High-end: Focus on performance, availability
 - Low-end: Focus on low cost, high capacity

Disk Interfaces



- Moving from Parallel to Serial:

| | |
|--------------------|--|
| SCSI-2 Ultra320 | Fibre Channel Serial Attached SCSI (SAS) |
| IDE, EIDE, ATA | Serial ATA (SATA) Fiber Attached Technology Adapted (FATA) |

Disk Interfaces: Parallel to Serial



- Technology Drivers:
 - Packages getting smaller:
 - Existing connectors and cables too big
 - Form factor: 3½ inch toward 2½ inch
 - Laptops and desktops shrinking
 - Blades
 - Higher data rates
 - Desire for lower power consumption
 - Lower logic voltages:
 - SCSI and ATA designed in 5V TTL era

Serial Technology Benefits



| Technology Feature | System Benefit |
|---|---|
| Embedded clock | Ease of design: No skew |
| Fewer signals | Simplified backplane routing |
| Thin cables | Smaller, lower cost, more flexible cables, better airflow |
| Point-to-point connections instead of bus-based | Dedicated bandwidth, scalable throughput |
| Performance beyond ATA & SCSI | Improves overall bandwidth |

Serial ATA (SATA)



- Next evolution of the ATA interface
 - Combines:
 - Serial connection
 - ATA (physical disk characteristics)
 - Focus: Low cost, high capacity, limited duty cycle

Serial Attached SCSI (SAS)



- Next evolution of the SCSI interface
 - There will be no Ultra640
 - Combines:
 - Parallel SCSI (command set)
 - Fibre Channel (frame formats)
 - SATA (physical characteristics)
 - Focus: High transaction rates and high reliability in enterprise environment

Fiber Attached Technology Adapted (FATA)



- Combines:
 - FC-AL interface for performance, availability
 - SATA mechanics for low cost
- Cost and performance between SATA and FC drives

Interface Positioning



- Serial ATA:
 - Low cost
 - ATA replacement
 - Single drive, single-user focus
- Serial Attached SCSI (SAS)
 - Private, local peripheral attachment
 - Parallel SCSI replacement
 - Highly Scaleable
- Fibre Channel
 - Very high performance or large scale attachment
 - Storage Area Networks (SANs)

The Battle in Storage Networking: Fibre Channel vs. Ethernet



- Both sides started by disparaging the other:
 - Ethernet folks said: Fibre Channel was an attempt at networking done by storage folks
 - FC folks said: iSCSI was an attempt at storage done by networking folks
- But each is now producing products for both worlds:
 - Cisco is doing Fibre Channel products: e.g. MDS 9000
 - Brocade is doing a Multi-Protocol Router which includes iSCSI functionality on Gigabit Ethernet

The Battle in Storage Networking: Fibre Channel vs. Ethernet



- There are presently some obstacles to iSCSI adoption:
 - Performance - TCP/IP protocol overhead is high
 - One solution: “iiSCSI HBA”
 - Fibre Channel installed base - Investment Protection
 - 10-gigabit Ethernet is expensive; 2 gigabit FC outperforms 1-gigabit Ethernet
- But also several favorable factors:
 - 1-gigabit Ethernet is less expensive than Fibre Channel
 - Ethernet and IP infrastructure is ubiquitous and well-understood

The Battle in Storage Networking: Fibre Channel vs. Ethernet



- Interoperability is also available, and improving:
 - FCIP (FC packets encapsulated in IP packets)
 - IP over FC (RFC 2625, IP packets sent over FC)
 - Cisco VSANs, and HP IP Storage Router
 - Brocade LSANs, FC Routing, and iSCSI gateway:
Multiprotocol Router
- Bottom line: Expect a long period of coexistence

- VSANs from Cisco
- Generally, as VLANs are to LANs, VSANs are to SANs
- FC port may be assigned to a VSAN
- Multiple VSANs can be configured, all independent
- Fault isolation is provided between different VSANs
- Cisco can also provide access between devices across VSANs with Inter-VSAN Routing

LSANs and Fibre Channel Routing



- An LSAN looks and is configured much like a Zone which can span multiple SAN fabrics
- Provides access to FC devices between SAN “islands”
- Fault isolation is provided between fabrics
- Fabric parameters may be different

Fibre Channel over Wide Area: SAN Extension

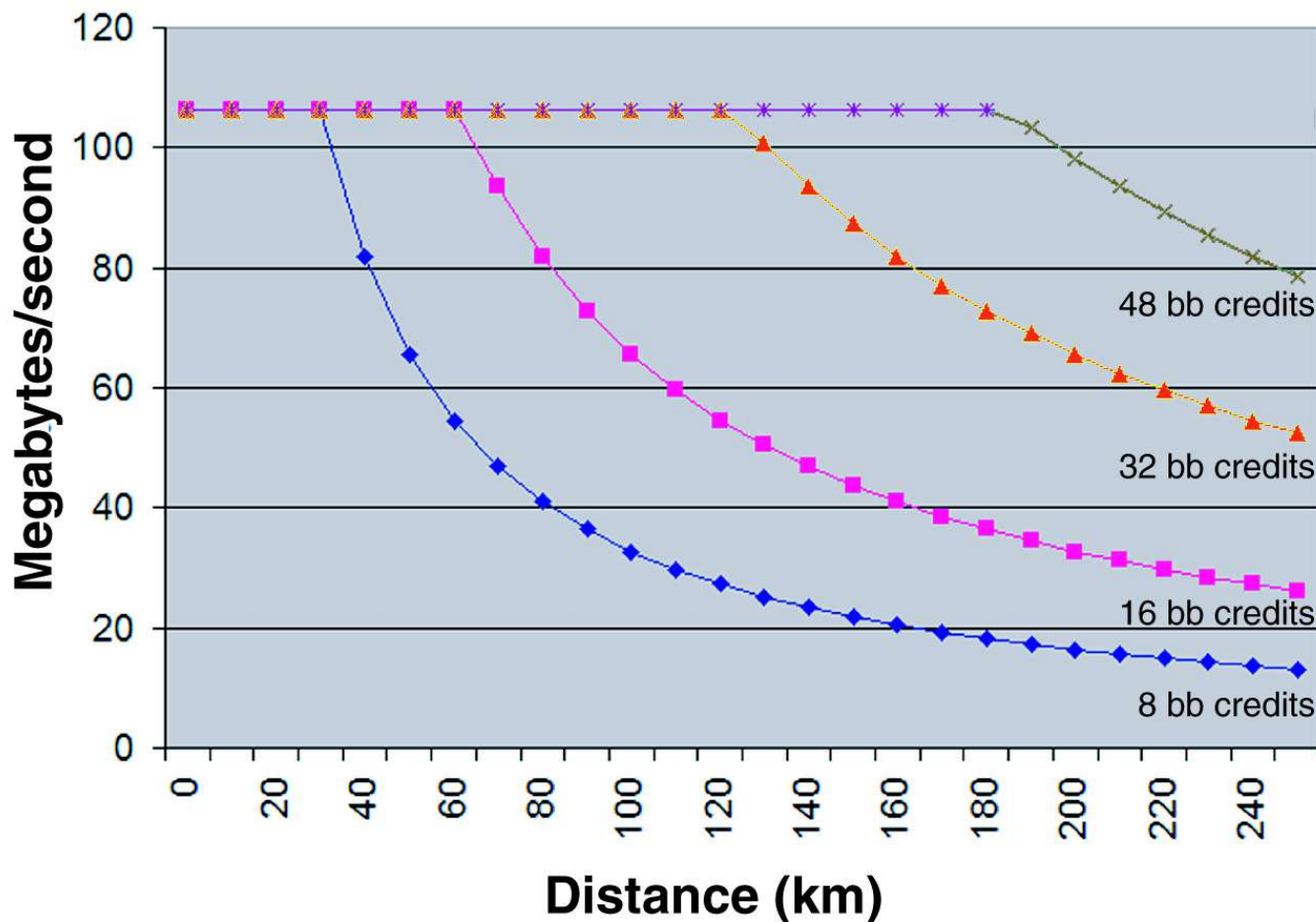


- Long-distance optical transceivers (GBICs, SFPs)
- Wave Division Multiplexing (WDM)
- Fibre Channel over IP (FCIP)
- Fibre Channel over SONET (FC-SONET)

Fibre Channel over Wide Area: Distance and the Effect of Buffer-to-Buffer Credits on Performance



Maximum FC Throughput (at 1 Gbps)



Major Storage Pain Points



- Survey of 211 sites running mission critical database applications:
 - #1 Pain Point: Managing Disk Space (40%)
 - #2 Pain Point: Backup (38%)
 - #3 Pain Point: Running out of Disk Space (31%)

Source: Richie Lary, TuteLary, LLC

Storage Controller Trends



- Virtualization
 - e.g. EVA "like an automatic camera compared with a manual one"
- Snapshot / Clone / Business Copy
- Controller-to-controller [remote] mirroring
 - e.g. HP Continuous Access, EMC SRDF
 - Disaster Recovery / Disaster Tolerance

- Virtualization makes automated provisioning possible, because of its flexible physical-to-logical mapping
- But something needs to actually do the provisioning, i.e. storage management software
- SMI-S (aka BlueFin) is a standard management protocol for storage systems, developed through Storage Network Industry Association (SNIA)
- SMI-S specifies how discovery, interrogation, and management of storage systems is done over Ethernet
- SMI-S has been widely accepted by vendors

Remote DMA (RDMA)



- Proprietary RDMA busses have been around for years
 - Digital CI, Tandem ServerNet, HP HyperFabric, Myrinet
- Infiniband was intended to be “the” open RDMA bus
 - but has only caught on in niches so far
- RDMA Consortium formed in 2002
 - Charter: build a standard transport for RDMA over TCP/IP
 - Included all major server/storage companies
 - HP and Compaq were active, founding members
 - Set of protocols known as iWARP announced in 2003
 - See <http://www.rdmaconsortium.org> for more details
 - RNICs being developed by many NIC vendors



i n v e n t

Speaker Contact Info



- Keith Parris
- E-mail: keith.parris@hp.com
- Web: <http://www2.openvms.org/kparris/>